Causal understanding is not necessary for the improvement of culturally evolving technology

Authors: Maxime Derex[1,2]*, Jean-François Bonnefon[3], Robert Boyd[4,5], Alex Mesoudi[1]

[1] Human Behaviour and Cultural Evolution Group, Department of Biosciences, University of Exeter, Penryn TR10 9FE, United Kingdom.
[2] Laboratory for Experimental Anthropology – ETHICS (EA 7446), Catholic University of Lille, 59016 Lille, France.
[3] Toulouse School of Economics (TSM-Research), Centre National de la Recherche Scientifique, University of Toulouse Capitole, Toulouse 31015, France.
[4] Institute of Human Origins, Arizona State University, Tempe, AZ 85287, USA.
[5] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287, USA.

*Correspondence to: maxime.derex@gmail.com

Highly-optimized tools are common in traditional populations. Bows and arrows, dogsleds, clothing, houses, and kayaks are just a few examples of the complex, exquisitely designed tools that humans produced and used to colonize new, demanding environments [1,2]. Because there is much evidence that humans' cognitive abilities are unparalleled [3,4], many believe that such technologies resulted from our superior causal reasoning abilities alone [5-7]. However, others have stressed that the high dimensionality of human technologies make them very hard to understand causally [8]. Instead, they argue that optimized technologies emerge through the selective retention of small improvements across generations without requiring explicit understanding of how these technologies work [1,9]. Here, we find experimental support for the latter view by showing that a physical artifact becomes progressively optimized across generations of social learners in the absence of explicit causal understanding. Moreover, we show that the transmission of causal models across generations has no noticeable effect on the pace of cultural accumulation. The reason is that participants do not spontaneously create multidimensional causal theories but instead mainly produce simplistic models related to a specifically salient dimension. Finally, we show that the transmission of these inaccurate theories 1) constrains exploration in subsequent generations of learners and 2) has negative downstream effects on their understanding. These results indicate that highly optimized technologies need not result from enhanced causal reasoning but instead can emerge from the accumulation of many small improvements made across generations linked by cultural transmission, and demand a focus on the cultural dynamics underlying technological change as well as individual cognition.

According to the *cognitive niche* hypothesis, natural selection enhanced our ancestors' ability to think creatively, plan and engage in causal reasoning about their environment [5,6], and these enhancements enabled the production of more efficient technologies that powered human expansion [10,11]. Our remarkable reasoning abilities certainly contribute to the development of sophisticated technologies [12]. Yet, others have stressed that even in traditional societies human technology is often too complex to be the product of human ingenuity *alone* [8,9]. Constructing a well-designed bow, for example, requires solving a difficult multi-dimensional optimization problem [13]. The *cultural niche* hypothesis suggests that complex technologies like bows result from the accumulation of many, mostly small, often poorly understood improvements made across generations linked by cultural transmission [1,9,14]. Over time, the selective retention of improvements gives rise to highly optimized solutions in the absence of explicit understanding about how these solutions work.

To test the hypothesis that the selective retention of beneficial changes over generations can produce cultural adaptations without individual understanding, we asked successive 'cultural generations' of participants (French university students) to optimize a physical system and measured participants' understanding of how the device worked at each generation (Fig. 1). The physical system was a wheel that traveled down a 1-meter long inclined track. The wheel had 4 radial spokes, and one weight could be moved along each spoke. Participants were organized into chains of 5 individuals. Each participant had 5 trials to minimize the time it took for the wheel to reach the end of the track. All participants (except those in the first generation) were provided with the last two configurations and associated scores of the previous participant in their chain so as to simulate overlapping generations. Participants were informed that their last two trials would be transmitted to the next participant in the chain, and that their reward depended both on their own performance and on the performance of the next participant in the chain. We collected data from 14 chains of 5 participants in this "Configurations" treatment.

The wheel system we used in this experiment suits our purpose for several reasons. First, it is unfamiliar (cognitive studies show that western students have poor understanding of wheel dynamics [15]), so participants cannot rely on acquired knowledge to solve the task. Second, the performance of the wheel depends solely on the laws of physics, and not on arbitrary principles that could compromise the ecological validity of our results. Finally, although the physics of the system are by no means trivial, the optimization problem is low-dimensional, which provides a conservative test of our hypotheses, compared to the many-dimensional problem of optimizing, for example, the performance of a bow [13].

The time required for the wheel to cover the track depends on just two variables: its moment of inertia and its initial potential energy (see Methods). This allowed us to rigorously measure participants' causal understanding of the system after they completed their 5 trials. Participants' understanding was evaluated by presenting them with pairs of wheels that differed in their configurations, and asking them to predict which wheel would reach the bottom of the rails first. A participant who understands the effects of varying the moment of inertia should predict that a wheel with 4 weights close to the axis would cover the track quicker than a wheel with 4 weights farther from the axis (Fig. 1A and B). Similarly, a participant who understands the role of potential energy should make correct predictions about the configurations displayed in Fig. 1C and D. The test comprised 10 pairs of wheels: 5 in which wheels varied in their moment of inertia, 5 in which wheels varied in their level of initial potential energy.

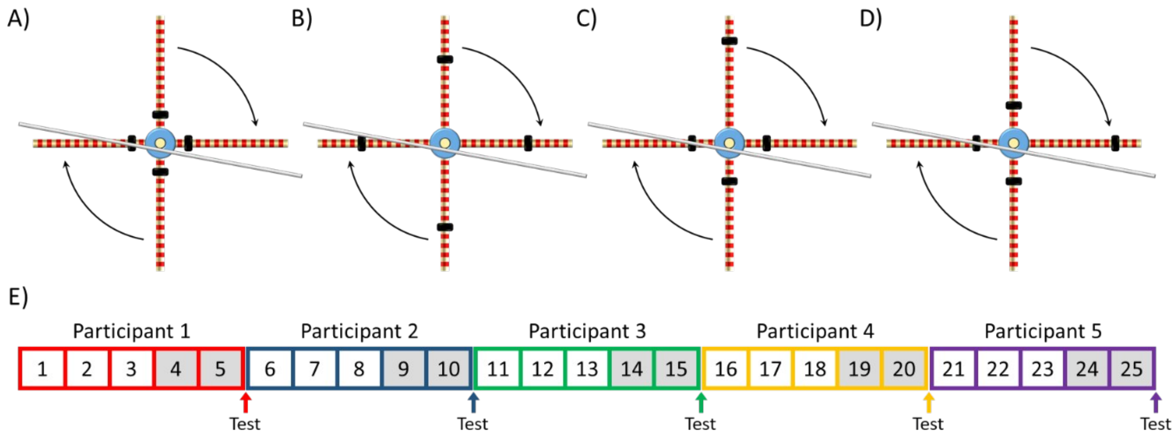

Figure 1 | Experimental task and design. A) Illustration of the physical system used in the experiment. The wheel had 4 radial spokes, and one weight could be moved along each spoke. The time it takes for the wheel to cover the track was determined by its moment of inertia and initial potential energy. A-B) The moment of inertia depends on how mass is distributed around the axis. Wheel A has a smaller moment of inertia and spins faster than wheel B. C-D) The amount of stored potential energy depends on the distance between the wheel centre of mass and the ground. Wheel C covers the distance faster than wheel D due to the higher initial position of its centre of mass. E) Participants were organized into chains of 5 individuals and had 5 trials each to improve their wheel. All participants (except those in the first generation) were provided with the last two configurations (shaded grey) and associated scores of the previous participant in the chain ("Configurations" treatment). Participants' understanding was evaluated after they completed their 5 trials by asking them to predict which of two wheels would cover the distance faster (e.g. A versus B, or C versus D).

The *cultural niche* hypothesis predicts that the speed of the wheel will increase with generations, while participants' understanding of the system will not improve over generations (preregistered hypothesis 1).

Results confirm these predictions. The average wheel speed (calculated as 1m/descent time) increased across generations (Generation 95% Confidence Intervals: (1.58, 9.02), mean =

5.37 m/h, Figure 2.A) while participants' understanding did not (Generation 95% CI: (-0.34, 0.25), mean = -0.04, Figure 2.B). The average wheel speed produced by first generation participants on their last trial was 123.6 m/h (95% Highest Posterior Density Interval: (117.3, 130.6)) and their understanding score was 4.60 (95% HPDI: (3.83, 5.53)). After 5 generations, average wheel speed increased to 145.7 m/h (95% HPDI: (138.5, 152.4)) while participants' understanding remained the same (95% HPDI: (3.65, 5.39), mean = 4.47). Given that the maximum possible speed was about 154 m/h, these results indicate an optimization of 71% after only four cultural generations. This confirms that the retention of improvements over generations produces highly optimized solutions and need not depend on the emergence of more accurate causal models.

To further investigate the relationship between cultural accumulation and individual understanding, we ran a second "Configurations + Theory" treatment with another 14 chains of 5 participants, in which participants could also formulate an explicit written theory about the physical system and transmit it to the next participant in the chain. The cultural transmission of explicit causal theories might affect both the optimization and the understanding of the physical system (preregistered hypothesis 2). One possibility is that theory transmission increases both individual understanding and wheel performance. For example, participants who have a correct representation of the wheel dynamics might enhance others' performance by helping them notice the effects of varying specific parameters. The effects of theory transmission, however, depend on the probability that participants generate useful theories. If participants produce incorrect theories, theory transmission would prevent individuals from noticing relevant parameters and detrimentally affect their performance. Inheriting a theory can also constrain participants' exploration behavior (preregistered hypothesis 3). For example, cognitive scientists have shown that children who are told the function of a toy engage in more limited exploration and are less likely to discover alternative functions than children ignorant of the toy's function [16], see also [17]. In our experiment, theory transmission might shape the exploration of parameter space and have negative downstream effects on participants' performance.

Results show that the average wheel speed increased at a similar rate in the "Configurations + Theory" as it did in the "Configurations" treatment (Treatment 95% CI: (-10.76, 18.13), mean = 3.52 m/h; Generation x Treatment 95% CI: (-7.07, 2.52), mean = -2.23 m/

h, Figure 2.A) and that participants' understanding again barely changed across generations, although participants in the very last generation had a slightly better understanding when they had inherited a theory (Treatment 95% CI: (-2.54, 0.31), mean = -1.14; Generation x Treatment 95% CI: (0.03, 0.81), mean = 0.44, Figure 2.B). Thus, these analyses do not provide substantial support for the idea that the transmission of explicit causal theories affects wheel optimization and individual understanding.

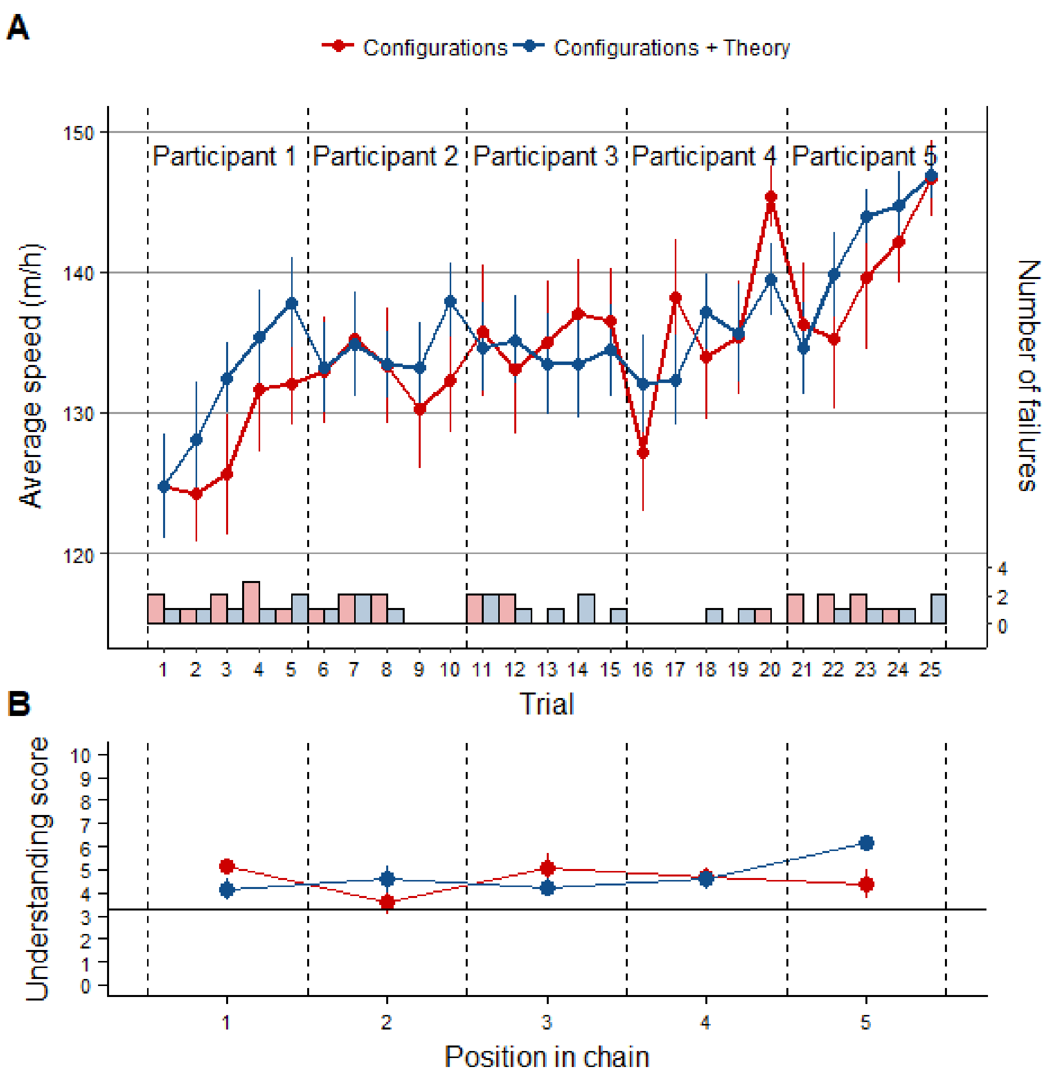

Figure 2 | Participants produce faster wheels across generations but their understanding of the system does not increase. A) Wheel performance across trials in the "Configurations" treatment (red bars and line) and "Configurations + Theory" treatment (blue bars and line). Vertical bars show the number of wheels that did not descend (i.e. failures) at each trial in each treatment. Coloured lines show the average speed for non-failure wheels at each trial in each treatment. B) Participants' understanding score across generations in each treatment. Horizontal line shows expected score for random guessers. Error bars show s.e.m.

Exploratory analyses, however, reveal striking differences between treatments in participants' exploration behavior (Fig. S7). To investigate the effect of theory transmission, participants' theories were coded according to whether they contained information related to moment of inertia, information related to potential energy, both, or neither. Of the 56 participants who inherited a theory (all participants in the "Configurations + Theory" treatment except first-generation participants), 15 inherited an inertia-related theory, 17 inherited an energy-related theory, 6 inherited a full theory and 18 inherited diverse, irrelevant theories. Participants who inherited an inertia theory mainly produced compact and balanced wheels (i.e. with low moment of inertia, Fig. 3B and F). In contrast, participants who inherited a potential energy theory produced unbalanced wheels with their top and right weights at extreme positions (i.e. with more energy and higher initial acceleration, Fig. 3C and G). The few participants who inherited a full theory produced compact and asymmetrical wheels (Fig. 3D and H). For comparison purposes, participants in the "Configurations" treatment (who did not inherit any theory) generated a greater range of wheels, although their center of mass tended to be concentrated in the upper-right quadrant (Fig. 3A and E).

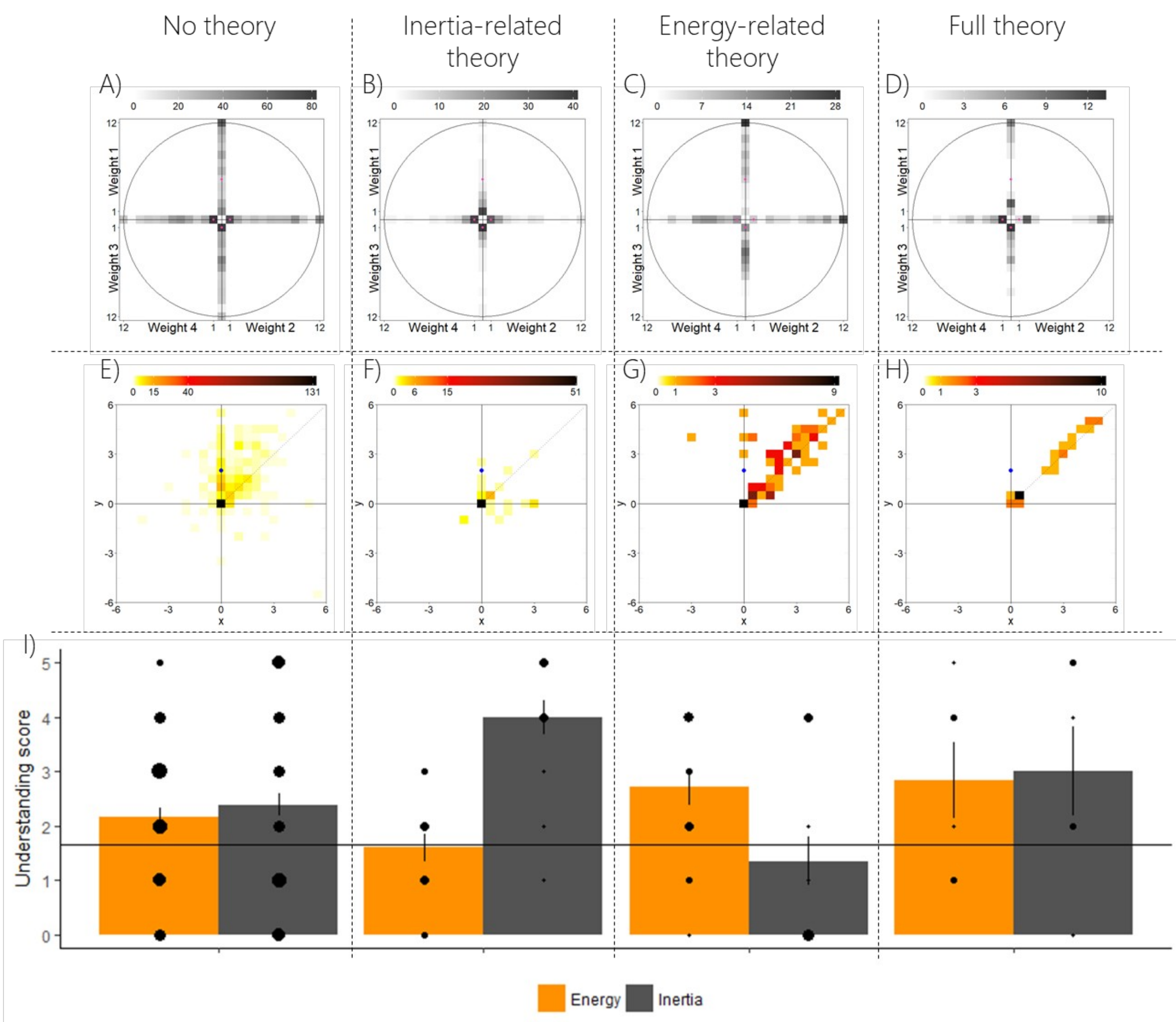

Figure 3 | Inheriting a theory affects both participants' exploration and understanding. A-D) Heat maps illustrating the most frequent weights' positions along each spoke. E-H) Heat maps illustrating the most frequent positions of the wheels' centres of mass (blue dot shows the optimum centre of mass position). Participants who did not inherit any theory sampled various positions along each spoke (A) and their wheels' centres of mass were concentrated in the upper-right quadrant (E). Participants who inherited an inertia theory mainly produced compact and balanced wheels (B-F). Participants who inherited a potential energy theory produced unbalanced wheels with their top and right weights at extreme positions (C-G). The few participants who inherited a full theory produced compact and asymmetrical wheels (D-H). Inheriting an inertia theory reduces understanding about energy and increases understanding about inertia, while inheriting an energy theory increases understanding about energy and reduces understanding about inertia (I). Horizontal line shows expected score for random guessers. Error bars show s.e.m. Black dots represent raw data with dot size representing the number of observations (I).

Furthermore, inherited theories strongly affected participant's understanding of the wheel system. Participants who did not inherit any theory ("Configurations" treatment) scored similarly (and better than chance) on questions about inertia and questions about energy (Fig. 3I). In comparison, participants who inherited an inertia- or energy- related theory showed skewed understanding patterns. Inheriting an inertia-related theory increased their understanding of inertia, but decreased their understanding of energy; symmetrically, inheriting an energy-related theory increased their understanding of energy, but decreased their understanding about inertia. One explanation for this pattern is that inheriting a unidimensional theory makes individuals focus on the effect of one parameter while blinding them to the effects of others. However, participants' understanding may also result from different exploration patterns. For instance, participants who received an inertia-related theory mainly produced balanced wheels (Fig. 3F), which could have prevented them from observing the effect of varying the position of the wheel's center of mass. To test this mechanism, we grouped participants who did not inherit any theory (i.e. from the "Configurations" treatment) into 3 categories: those who produced various types of wheels, those who only produced balanced wheels, and those who only produced unbalanced wheels. Participants who produced various types of wheels scored similarly on questions about inertia and energy. However, participants who only produced balanced wheels showed better understanding of inertia than energy, and participants who only produced unbalanced wheels showed better understanding of energy than inertia (Fig. S8). These results suggest that the understanding patterns observed in participants who received unidimensional theories is likely the result of the canalizing effect of theory transmission on exploration. Note that in the present case, this canalizing effect is performance-neutral: with our 2-dimensional problem, better understanding of one dimension and worse understanding of one dimension simply compensate each other. For a many-dimensional problem, though, better understanding of one dimension is unlikely to compensate for worse understanding of all the others.

As predicted by the *cultural niche* hypothesis [9], our experiment shows that highly optimized technologies can emerge from the accumulation of many improvements made across generations linked by cultural transmission, without the need for an accurate causal understanding of the system. Most participants actually produced incorrect or incomplete theories despite the relative simplicity of the physical system. These results are consistent with the view that individuals do not spontaneously create multidimensional representations of object motion [15]. Instead they mainly produce unidimensional models related to a specifically salient dimension [18]. Although evidence of individuals' erroneous theories of motion are sometimes considered as experimental artefacts resulting from impoverished stimuli (such as using pictures to describe dynamical events; [19]), our results show that incomplete representations commonly emerge even when individuals directly observe and modify an actual physical object. As a consequence, the transmission of explicit theories across generations did not help participants produce more efficient wheels: inheriting a theory mostly constrained participants' exploration, and prevented them from noticing the effects of relevant variables outside of the theory they received.

It is worth noting that despite exhibiting poor understanding of the experimental physical system, participants did not randomly explore the parameter space. For example, in both treatments, wheels were much more likely to have their center of mass at the center of the wheel, or in the upper right quadrant. This indicates that participants had appropriate intuitions about how to maximize acceleration, and sampled the parameter space fairly efficiently in that regard. Our ability to restrict exploration to potentially useful portions of the design space certainly accelerated cultural evolution in our experiment. A greater focus on the determinants of biased exploration would be a fruitful area for further work. Here, we cannot tell whether participants' intuitions resulted from an implicit physics engine, from past experience with analogous objects, or from western formal education (although physics or engineering background had no effect on participants' understanding scores, Fig. S9). Future cross-cultural work involving non WEIRD participants should tell us whether this selective exploration is culturally constructed or shared across populations [20]. In any case, our experiment indicates that one should be cautious when interpreting complex archaeological materials as evidence for sophisticated cognitive abilities (such as reasoning, problem-solving or planning), since these abilities are not the sole driver of technological sophistication [1,9]. Understanding technological change demands a focus on individual cognition [5,6] but also requires to give attention to factors affecting the pace of cultural accumulation, such as cultural transmission dynamics and demography [21-29].

Methods:

Experimental apparatus

Dynamics of the wheel

The performance of the wheel depends on two variables: its moment of inertia and its initial potential energy. The wheel's moment of inertia depends on how mass is distributed around its axis of rotation. Wheels with a smaller moment of inertia (i.e. wheels that have their weights closer to the axis) require less torque to increase angular momentum and spin faster (see Movie S1 and S2). The amount of potential energy stored in the wheel depends on the distance between the wheel's centre of mass and the ground at its initial position (see Movie S3 and S4). When the centre of mass of the wheel is in the wheel's upper right quadrant, more potential energy is converted into angular kinetic energy so that the wheel will benefit from higher increases in angular momentum. Note that the same would occur with a center of mass in the upper left quadrant. There, the wheel would rotate in the wrong direction and would go up on the rails (the kinetic energy would be converted back into potential energy).

In our experiment, both the wheel's moment of inertia and its potential energy had to be taken into account to reach the best performance. Potential energy could not be stored without increasing the wheel's moment of inertia and so there was a tradeoff between storing energy and minimizing inertia (Fig. S1). Potential energy could be efficiently exploited in two different ways. One is keeping all weights close to the axis except the top one. The other is moving both the top and right weights away from the axis. This latter strategy can give the wheel better initial acceleration because the right weight has more leverage than the top weight to set the wheel in motion at its initial position (the top weight initially applies a vertical force on the axis which doesn't affect the wheel's angular momentum). However, the right weight will only fall from half the height of the top weight (assuming both weights are equally far from the axis) so less potential energy will eventually be converted into kinetic energy.

Building of the wheel

The wheel was built around a tube clamp designed to form a 90-degree angle between a 28 mm tube (which passed through the clamp) and four other 28 mm tubes (with 90-degree angles between contiguous tubes, see Fig. 1 and S2A). The axis of the wheel was composed of a 10.5 cm long bored-through wooden pole and an 8 mm threaded steel rod in its centre. The threaded steel rod protruded approximately 4 cm past the end of the wooden pole at each side and was covered with pieces of 3 cm rubber tube in order to prevent the wheel from sliding on the rails. Flat washers were positioned on either side of the pieces of rubber tube to guide the wheel along the rails and limit potential friction. Two nuts held the materials in position. Two 500-gram weight plates were positioned along the axis of the wheel (one on each side of the clamp) in order to reduce the wheel's moment of inertia and limit the occurrence of motionless or back-spinning configurations. Two barbell clamp collar clips were used to lock the weight plates in position (Fig. S2B). Four 28 mm wooden poles formed the spokes of the wheel and were 41 cm long from the centre of the wheel. Pieces of red tape were positioned every 28 mm along the spokes in order to signal 12 discrete weights' potential positions (closest position to the axis was 6.5 cm from the centre of the wheel). Four barbell clamp collar clips were used as weights. Each was weighted with flat washers, screws and nuts (Fig. S2C). The weight of a collar clip was about 100 grams.

Building of the rails

Rails were built from 2 meter long plated steel slotted angles (20mm wide). A steel and aluminium structure held the rails at an incline of 14 degrees. Two push-button switches (made from computer mice) were located 92 cm apart on the rails and connected to a computer program (Fig. S2B). Two arrows indicated the positions of the switches (starting/ending points, Fig. S2A). A mechanical lever maintained the wheel motionless, with 2 of its spokes parallel to the ground at its starting position.

<u>Participants</u>

In total, 140 participants took part in the study (70 women and 70 men). Participants were randomly selected from a database managed by Catholic University of Lille and recruited by email from various universities in Lille, France. The subjects ranged in age from 18 to 38 y (mean of 20.5, SD of 3.4). Participants received 3€ for participating and an additional amount ranging from 0 and 26€ depending on their own performance and the performance of the next participant in their chain (see below).

<u>Ethical statement</u>

The study was carried out in accordance with the ethical standards of the 1964 Declaration of Helsinki and the guidelines of the British Psychological Society's Code of Human Research Ethics. All methods were approved by the University of Exeter Biosciences Research Ethics Committee (2018/2310) and the Catholic University of Lille Research Ethics Committee (2018-01-31-E). All participants provided written, informed consent before taking part in the experiment.

Procedure

The experiment took place in an experimental room at the Laboratory for Experimental Anthropology at Catholic University of Lille. For each session (around 20 minutes long), a single individual was recruited and sat at a computer that was placed parallel to and at 2 meters from the experimental apparatus. Participants were randomly assigned to one condition of the experiment and one sex-segregated chain. Before starting the experiment, participants were asked to complete a consent form and were asked their age. At the end of the experiment, participants indicated whether they have an academic background in physics or engineering. Participants entered and left the room by two different doors to prevent any form of direct interactions between participants. Participants came back to the lab a few days after the experiment to get paid (once their final payoff was known, see below).

Experimental design

Building phase

Each participant had 5 trials to minimize the time it took the wheel to cover about one meter on an inclined track. Weights could be placed on one of 12 discrete positions along 4 spokes which created a space of 20,736 unique configurations. Participants chose their configurations through a computer program using 4 sliders (Fig. S3 and Computer program S1). Once the configuration was confirmed by the participant, the experimenter positioned the weights on the physical wheel accordingly (the computer screen was projected onto a wall to the right of the participant in order to allow the experimenter to see the chosen configuration without interacting with the participant, Fig. S2A). The wheel was then positioned on the rails and held motionless by a mechanical system before being released. Once released, the time it took the wheel to descend the track was automatically recorded by the computer program. The wheel's average speed and associated payoff was then automatically displayed on the participant's screen. Participants could consult their two last configurations between any trials. They had as much time as they needed to consult these configurations and choose their next one. After 3 trials, participants were reminded that their last two configurations will be transmitted to the next participant in the chain. After five trials, the program automatically switched to the test phase.

Testing phase

After completing the task, participants were told that they would be presented with pairs of wheels and that they must guess which one of 2 wheels would descend the rails faster. They were also told that one of their answers will be randomly selected at the end of the test and that 5 euros will be added to their gain if that answer is correct. For each pair, participants could submit 3 possible answers: “Wheel 1”, “Wheel 2” or “No difference”. Participants could take as much time as needed before submitting their answer. Once an answer was submitted, another pair of wheels was displayed until participants compared 10 pairs of wheels. In 5 pairs, wheels varied in their moment of inertia, in the other 5, wheels varied in their level of initial potential energy (Fig. S4). Participants were not told whether their guesses were correct. All participants were exposed to the same 10 pairs of wheels in the same order.

Experimental treatments

Two treatments were run. In each treatment, participants were part of 14 chains each containing 5 individuals (exclusively males or exclusively females). All participants except those in the first generation were provided with social information. In the “Configurations” treatment (n = 70), the last two configurations and associated scores of the previous participant in the chain were provided to the next participant in the chain. In the “Configurations + Theory” treatment (n =70), participants additionally received the previous participant’s theory about the physical system. Participants were asked to write their theory after the test phase was completed. Participants could not transmit information about the performance of a specific configuration in order to prevent individuals from extending the number of transmitted configurations as compared to the “Configurations” treatment. Theories had to be less than 340 characters long and always started with “The wheel covers the distance faster when...”. Social information was available all along the building phase and could be consulted between any trials in both treatments.

Pre-experiment information

Instructions could be read on a computer screen and stated that the participants’ task was to position 4 weights on a wheel in order to minimize the time it takes the wheel to cover an inclined track (see Computer program S1). Participants were informed that they have 5 trials to do this and that their payoff will be determined by the performance of each of their wheels. Participants were told that they were part of a chain and so that the task was a collective one (despite being alone in the experimental room). They were informed that their last two configurations will be transmitted to the next participant in the chain and all participants except those in the first generation were also told that they were going to be provided with the last two configurations of the previous participant in the chain. In the “Configurations + Theory” treatment, participants were also informed that they could write/receive a theory. Finally, participants were told that their final gain will be determined by their own performance and the performance of the next participant in the chain. Participants did not know the length of the chain nor the speed of the best possible wheel.

Participants' payoff

The following equation determined the payoff of each wheel:

$$[1 - ((\text{MaxSpeed} - \text{RecordedSpeed}) / (\text{MaxSpeed} - \text{MinSpeed}))] \times 3 + \text{Bonus}$$

with MaxSpeed = 160, MinSpeed = 96. RecordedSpeed was the recorded average speed of the wheel. Bonus took the value 0.2 for wheels that descended and 0 otherwise.

Participants' final payoff corresponded to the sum of the payoff of each of their wheels plus the payoff of the next participant' first two wheels plus 5€ if they correctly answered the randomly selected test. Final participants in chains had their last two payoffs doubled (although they were not aware of this as they didn't know that the chain was about to end).

Theory coding

5 individuals blind to the research question were explained the dynamics of the wheel (i.e. the respective role of inertia and energy in the performance of the wheel) and were asked to code participants' theories according to whether they contain accurate information related to moment of inertia and/or potential energy. A theory contained information related to moment of inertia when it says that the wheels goes faster when its weights are close to the axis (e.g. "*The wheel covers the track faster when its weights are balanced and close to the axis.*"). A theory contained information related to potential energy when it says that the wheel goes faster when its center of mass is in the upper-right quadrant (e.g. "*The wheel covers the track faster when its top and right weights are farther from the axis than its bottom and left weights.*"). A few theories contained information about both principles (e.g. "*The wheel covers the track faster when its weights are balanced and close to the axis. Furthermore the wheel has a better initial acceleration when the top and rights weights are slightly farther away from the axis.*"). Cohen's kappa coefficients reveal almost perfect agreement between raters (0.81 for inertia and 0.85 for energy).

<u>Statistical analyses and models output</u>

We ran a series of Bayesian multi-level models in R [30]. Models were fitted using map2stan in the *rethinking* package [31] and 95% credible intervals were used to make inferences.

Analysis 1

Preregistered analysis 1 investigated the average speed of wheels across generations in the Configurations treatment. Wheels that did not go down were attributed a speed of 0. Data were restricted to participants' last two trials in order to limit the occurrence of wheels that did not descend in the dataset. We fitted a linear model with "Speed" as the outcome variable, "Trial", "Generation" as predictor variables and "Player's identity" and "Chain's identity" as random effects (see Table S1 for model output).

Analysis 2

Preregistered analysis 2 investigated understanding across generations in the Configurations treatment. We fitted a linear model with "Score" as the outcome variable, "Generation" as a predictor variable and "Chain's identity" as a random effect (see Table S2 for model output).

Analysis 3

Preregistered analysis 3 compared the average speed of wheels across generations between treatments. Wheels that did not go down were attributed a speed of 0. Data were restricted to participants' last two trials in order to limit the occurrence of wheels that did not descend in the dataset. We fitted a linear model with "Speed" as the outcome variable, "Trial", "Generation", "Treatment", "Trial:Treatment" and "Generation:Treatment" as predictor variables and "Player's identity" and "Chain's identity" as random effects (see Table S3 for model output). For this model, the chains were inefficient and the effective number of samples for one parameter was low (Table S3). The robustness of the model estimates was checked by running additional models (see below). Additional models with more efficient sampling confirmed the reported results (supplementary analysis 1, Table S5 and S6).

Analysis 4

Preregistered analysis 4 compared understanding across generations between treatments. We fitted a linear model with "Score" as the outcome variable, "Generation", "Treatment" and "Generation:Treatment" as predictor variables and "Chain's identity" as a random effect (see Table S4 for model output).

Deviation from preregistered analyses

In preregistered analysis 4, the outcome variable was "Score" and each participant was associated with 2 values in the dataset: one score for inertia, the other for energy. As compared to the analysis we ran, the preregistered model included "Physical Principle" and "Physical Principle: Treatment" as predictor variables and "Player's identity" as random effect. However, analyses revealed that understanding scores about inertia and energy were negatively correlated (Fig. S6 and Table S7) and some individuals better understood inertia than energy while others better understood energy than inertia (Fig. 3I and S8). As a result, the preregistered model did not converge so we ran our analysis on aggregated score and removed the terms associated the variable “Physical Principle” in the reported model.

References:

1 Henrich, J. *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. (Princeton University Press, 2015).

2 Richerson, P. J. & Boyd, R. *Not by genes alone*. (University of Chicago Press, 2005).

3 Povinelli, D. J. *World Without Weight: Perspectives on an Alien Mind*. (Oxford University Press, 2011).

4 Reader, S. M. & Laland, K. N. Social intelligence, innovation, and enhanced brain size in primates. *Proc. Natl. Acad. Sci. U.S.A* **99**, 4436-4441, doi:10.1073/pnas.062041299 (2002).

5 Pinker, S. The cognitive niche: Coevolution of intelligence, sociality, and language. *Proc. Natl. Acad. Sci. U.S.A* **107**, 8993-8999 doi:10.1073/pnas.0914630107 (2010).

6 Barrett, H. C., Cosmides, L. & Tooby, J. The hominid entry into the cognitive niche. *Evolution of mind, fundamental questions and controversies*, 241-248 (2007).

7 Bingham, P. M. Human Uniqueness: A General Theory. *Q. Rev. Biol.* **74**, 133-169 (1999).

8 Boyd, R., Richerson, P. J. & Henrich, J. The cultural evolution of technology: facts and theories. *Cultural evolution: society, technology, language, and religion*, 119-142 (2013).

9 Boyd, R., Richerson, P. J. & Henrich, J. The cultural niche: Why social learning is essential for human adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10918-10925 (2011).

10 Kyriacou, A. & Bruner, E. Innovation and the Evolution of Human Behavior Brain Evolution, Innovation, and Endocranial Variations in Fossil Hominids. *PaleoAnthropology* **2011**, 130-143 (2011).

11 Fuentes, A. *The creative spark: How imagination made humans exceptional*. (Penguin, 2017).

12 Derex, M. & Boyd, R. The foundations of the human cultural niche. *Nat. Commun.* **6**, 8398 (2016).

13 Baker, T. Bow design and performance. *The traditional bower's Bible* **1**, 43-116 (1992).

14 Muthukrishna, M. & Henrich, J. Innovation in the collective brain. *Philos. Trans. R. Soc. B-Biol. Sci.* **371**, doi:10.1098/rstb.2015.0192 (2016).

15 Proffitt, D. R., Kaiser, M. K. & Whelan, S. M. Understanding wheel dynamics. *Cogn. Psychol.* **22**, 342-373, doi:https://doi.org/10.1016/0010-0285(90)90007-Q (1990).

16 Bonawitz, E. *et al.* The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition* **120**, 322-330, doi:http://dx.doi.org/10.1016/j.cognition.2010.10.001 (2011).

17 Wood, L. A., Kendal, R. L. & Flynn, E. G. Does a peer model's task proficiency influence children's solution choice and innovation? *J. Exp. Child. Psychol.* **139**, 190-202 (2015).

18 Proffitt, D. R. & Gilden, D. L. Understanding natural dynamics. *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 384 (1989).

19 Kubricht, J. R., Holyoak, K. J. & Lu, H. Intuitive Physics: Current Research and Controversies. *Trends Cogn. Sci.* **21**, 749-759, doi:10.1016/j.tics.2017.06.002 (2017).

20 Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61-83 (2010).

21 Henrich, J. Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses - The Tasmanian case. *Am. Antiq.* **69**, 197-214 (2004).

22 Powell, A., Shennan, S. & Thomas, M. G. Late Pleistocene Demography and the Appearance of Modern Human Behavior. *Science* **324**, 1298-1301 (2009).

23 Kline, M. A. & Boyd, R. Population size predicts technological complexity in Oceania. *Proc. R. Soc. B-Biol. Sci.* **277**, 2559-2564 (2010).

24 Derex, M., Beugin, M.-P., Godelle, B. & Raymond, M. Experimental evidence for the influence of group size on cultural complexity. *Nature* **503**, 389-391 (2013).

25 Muthukrishna, M., Shulman, B. W., Vasilescu, V. & Henrich, J. Sociality influences cultural complexity. *Proc. R. Soc. B-Biol. Sci.* **281**, 20132511, doi:10.1098/rspb.2013.2511 (2014).

26 Hill, K. R., Wood, B. M., Baggio, J., Hurtado, A. M. & Boyd, R. T. Hunter-Gatherer Inter-Band Interaction Rates: Implications for Cumulative Culture. *PLoS One* **9**, e102806, doi:10.1371/journal.pone.0102806 (2014).

27 Derex, M. & Boyd, R. Partial connectivity increases cultural accumulation within groups. *Proc. Natl. Acad. Sci. U.S.A* **113**, 2982-2987 (2016).

28 Creanza, N., Kolodny, O. & Feldman, M. W. Greater than the sum of its parts? Modelling population contact and interaction of cultural repertoires. *J. R. Soc. Interface* **14**, 20170171 (2017).

29 Derex, M., Perreault, C. & Boyd, R. Divide and conquer: intermediate levels of population fragmentation maximize cultural accumulation. *Philos. Trans. R. Soc. B-Biol. Sci.* **373**, 20170062 (2018).

30 R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2011).

31 McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. (CRC Press, 2016).

Acknowledgments. We thank J. Clewett for valuable advice about the building of the wheel, F. Gosselin and A. Deymier for organizing the experimental sessions, and members of the laboratory for experimental anthropology for helpful discussions during the development of the experimental protocol. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 748310.

Author contributions. R.B. and M.D. developed the research question. M.D. conceived the experimental task and protocol with input from A.M. and J.-F.B. M.D. performed the experiment, analyzed the data and wrote the manuscript with input from all authors. Authors declare no competing interests.

Competing interests. The authors declare no competing financial interests.

Materials & Correspondence. Correspondence should be addressed to M.D. (maxime.derex@gmail.com). Preregistered hypotheses and analyses are available at osf.io/ge7cs. Data and scripts are available at osf.io/afwmr.

Supplementary Materials. Additional Methods, Figures, Tables Movies and Source Data are available as supplementary materials.